





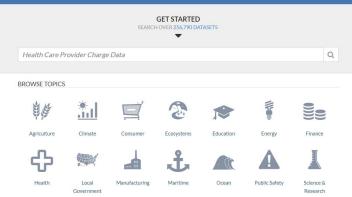
## 获取气象数据



DATA TOPICS - IMPACT APPLICATIONS DEVELOPERS CONTACT

#### The home of the U.S. Government's open data

Here you will find data, tools, and resources to conduct research, develop web and mobile applications, design data visualizations, and <u>more</u>.



#### 从www.data.gov下载气象数据

#### Global Historical Climatology Network - Daily (GHCN-Daily), Version 3

National Oceanic and Atmospheric Administration, Department of Commerce — The Global Historica Climatology Network - Daily (GHCN-Daily) dataset integrates daily climate observations from approximately 30 different data sources. Version 3...

HTML HTML TEXT TEXT PDF HTML 16 more in dataset

数据集整合了来自大约30个不同数据源的每日气候观测。 包含了全球90,000多个陆基台站的台站测量数据,从中 只取出中国站点的数据。

## 气象数据简介

获取的气象数据

## 例如气象数据文件CHM00045004.dly的主要内容:

CHM00045004199208TMAX 319 S 307 S 320 S 318 S 318 S 318 S 330 S .....

CHM00045004199208TMIN 268 S 268 S 274 S 276 S 277 S 279 S -9999 .....

CHM00045004199208PRCP 0 S 0 S 0 S 0 S 0 S 10 S 290 S .....

CHM00045004199208TAVG 291H S 287H S 295H S 295H S 294H S 296H S ......

注: 其中每行数据中红色为站点编号紫色表示年份,黄色表示月份,绿色表示气象数据类型。TMAX表示每天最高气温,TMIN表示每天最低气温,PRCP表示降水量,TAVG表示每日平均气温。

### 文件stations.txt主要记录了气象站点的信息

CHM00045004	22.3330	114.1670	26.0	KING'S PARK	45004
CHM00045005	22.3000	114.1670	62.0	HONG KONG OBSERVATORY	45005
CHM00050353	51.7170	126.6500	179.0	HUMA	50353
CHM00050434	50.4500	121.7000	733.0	TULIHE	50434

只利用其中的站点编号、经纬度和站点名称。

#### 名称

- CHM00045004.dly
  CHM00045005.dly
  CHM00050353.dly
  CHM00050434.dly
  CHM00050527.dly
  CHM00050557.dly
  CHM00050632.dly
  CHM00050632.dly
  CHM00050658.dly
- CHM00050727.dly
- CHM00050745.dly



## 用PySpark处理数据



### 将所有文件读为一个RDD

rdd = sc.wholeTextFiles("file:///" + os.getcwd() + "/china/")

### 数据清洗

去除字母, -9999等无效数据



### 进行计算、排序等操作

计算出最大值,最小值,平均值,并按照年份排序。最终得出各个站点以及全国每年的主要气象数据的统计结果。



将得出的结果和stations.txt文件中站点的主要 信息分别存入MySQL数据库中。



## 最终统计结果

id	TMAX	TMIN	TAVG	PRCP	SNWD
CHM000450041992	34.7	10.5	24.3	11.9	(Nul
CHM000450041993	32.9	5.1	22.8	5.6	(Nul
CHM000450041994	34.3	7	23.2	7.1	(Nul
CHM000450041995	34.1	-3.1	22.5	7.3	(Nul
CHM000450041996	34.4	5.3	22.8	6.1	(Nul
CHM000450041997	33.4	9.3	22.8	8.8	(Nul
CHM000450041998	34.2	8	23.4	7.4	(Nul
CHM000450041999	34.7	8.4	23.2	4	(Nul
CHM000450042000	27.7	-0.1	16.8	1.9	(Nul
CHM000450051946	33.9	9.4	22.6	5.4	(Nul
CHM000450051947	35	5.6	22	8	(Nu
CHM000450051948	33.3	3.9	22.7	8.5	(Nul
CHM000450051949	35.6	6.1	22.7	7.4	(Nul
CHM000450051950	33.9	6.1	22.4	6.9	(Nul
CHM000450051951	33.9	5	22.1	8.1	(Nul
CHM000450051952	33.3	4.4	22.7	8.1	(Nul
CHM000450051953	36.1	8.3	22.4	7.6	(Nul
CHM000450051954	35.6	6.1	22.6	5.2	(Nul
CHM000450051955	33.3	4.4	22.6	7.8	(Nul
CHM000450051956	35	3.9	22.4	5.5	(Nul
CHM000450051957	33.9	3.9	22.4	(Null)	(Nul
CHM000450051958	35	7.2	22.9	(Null)	(Nul
CHM000450051959	32.8	6.1	23.1	(Null)	(Nul

(Null)	2.3	11.8	-45.6	44.9	1998
(Null)	1.9	11.5	-43.3	46.2	1999
24.5	2.1	10.9	-45.1	47.7	2000
17	2.1	11.2	-49.6	47	2001
(Null)	2.1	11.5	-43	44.9	2002
(Null)	2	11.2	-45.6	43.1	2003
(Null)	1.8	11.4	-45.6	46.2	2004
(Null)	2.1	11	-41.2	45.4	2005
(Null)	2	11.6	-46.3	47.7	2006
(Null)	2	11.9	-41.1	46.3	2007
(Null)	2.2	11.3	-46.4	47.8	2008
(Null)	1.9	11.4	-47.5	44.7	2009
(Null)	2.3	11.4	-46.5	3276.6	2010
1	1.9	10.9	-47.4	47.8	2011
(Null)	2.3	10.7	-46.9	45.7	2012
7	4.2	11.5	-46.1	42.7	2013
6.2	4.2	11.6	-45.1	44.8	2014
4.7	5.1	11.6	-41.4	46.1	2015
4.3	5.7	11.7	-44.7	39.7	2016
4.6	5	11.9	-44.1	42.3	2017
6.1	5.5	11.7	-47.8	43.2	2018

d	latitude	longitude	name
CHM00045004	22.333	114.167	KING'S
CHM00045005	22.3	114.167	HONG
CHM00050353	51.717	126.65	HUMA
CHM00050434	50.45	121.7	TULIHE
CHM00050527	49.25	119.7	HAILAR
CHM00050557	49.167	125.233	NENJIANG
CHM00050564	49.433	127.35	SUNWU
CHM00050632	48.767	121.917	BUGT
CHM00050658	48.05	125.883	KESHAN
CHM00050727	47.167	119.933	ARXAN
CHM00050745	47.383	123.917	QIQIHAR
CHM00050756	47.433	126.967	HAILUN
CHM00050788	47.233	131.983	FUJIN
CHM00050854	46.383	125.317	ANDA
CHM00050915	45.517	116.967	ULIASTAI
CHM00050949	45.083	124.867	QIAN
CHM00050953	45.75	126.767	HARBIN
CHM00050963	45.967	128.733	TONGHE
CHM00050968	45.217	127.967	SHANGZHI
CHM00050978	45.283	130.949	JIXI

所有站点每年气象信息

全国每年气象信息

所有站点基本信息



## 数据可视化

主要运用了Python的第三方库matplotlib和pyecharts 首先从MySQL中读取数据,再选择合适的图表绘制。

# Matplotlib 简介

Matplotlib是一个Python 2D绘图库,它以多种硬拷贝格式和跨平台的交互式环境生成出版物质量的图形。 Matplotlib可用于Python脚本,Python和IPython以 Shell、Jupyter以笔记本,Web应用程序服务器和四个图形用户界面工具包。









Matplotlib 尝试使容易的事情变得更容易,使困难的事情变得可能。 您只需几行代码就可以生成图表、直方图、功率谱、条形图、误差图、散点图等。 更多的示例,请参见基础绘图例子和示例陈列馆。

为了简单绘图,该 pyplot 模块提供了类似于MATLAB的界面,尤其是与IPython结合使用时。 对于高级用户,您可以通过面向对象的界面或MATLAB用户熟悉的一组功能来完全控制线型,字体属性,轴属性等。

#### 简介

#### ☑ 概况

Echarts 是一个由百度开源的数据可视化,凭借着良好的交互性,精巧的图表设计,得到了众多开发者的认可。而Python 是一门富有表达力的语言,很适合用于数据处理。当数据分析遇上数据可视化时,pyecharts 诞生了。

### ♦ 特性

- 简洁的 API 设计,使用如丝滑般流畅,支持链式调用
- 襄括了30+种常见图表,应有尽有
- 支持主流 Notebook 环境,Jupyter Notebook 和 JupyterLab
- 可轻松集成至 Flask,Django 等主流 Web 框架
- 高度灵活的配置项,可轻松搭配出精美的图表
- 详细的文档和示例,帮助开发者更快的上手项目
- 多达 400+ 地图文件以及原生的百度地图, 为地理数据可视化提供强有力的支持

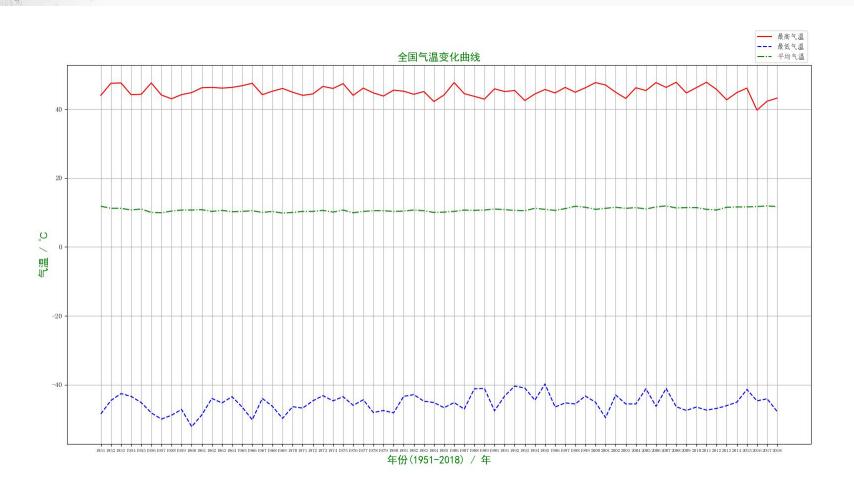
### ☑ 版本

pyecharts 分为 v0.5.X 和 v1 两个大版本,v0.5.X 和 v1 间不兼容,v1 是一个全新的版本,详见  $\underline{\text{ISSUE\#892}}$ , $\underline{\text{ISSUE\#1033}}$ 。

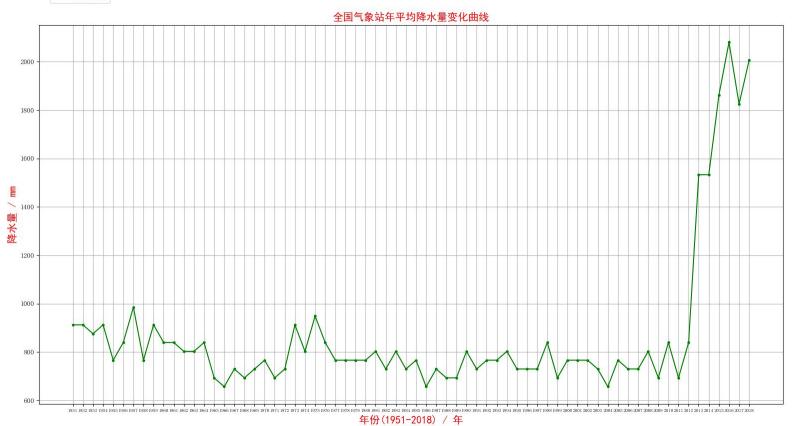
图片来自https://www.matplotlib.org.cn/intro/

图片来自https://pyecharts.org/#/zh-cn/intro

# 部分可视化效果展示







#### 全国2018年各气象站降水量分布图(单位/mm)



300-600 0-300

high

3200以上 2400-3200 1600-2400 1200-1600 600-1200

#### 全国2018年各气象站平均气温分布图(单位/℃)



-4-0



## 总结

